

CLAIMS

We claim:

- 1 1. A method of generating a synthetic voice comprising:
 - 2 receiving a user selection of a first text-to-speech (TTS) voice and a second TTS
 - 3 voice from a plurality of TTS voices;
 - 4 receiving at least one user-selected voice characteristic; and
 - 5 generating a new TTS voice by blending the first TTS voice and the second TTS
 - 6 voice and according to the user-selected voice characteristic.

- 1 2. The method of claim 1, further comprising:
 - 2 presenting the new TTS voice to the user for preview;
 - 3 receiving user-selected adjustments; and
 - 4 presenting a revised TTS voice to the user for preview according to the user-
 - 5 selected adjustments.

- 1 3. The method of claim 1, wherein generating the new TTS voice further comprises
 - 2 interpolating between corresponding segment parameters of the first TTS voice and the
 - 3 second TTS voice.

- 1 4. The method of claim 1, wherein the user-selected voice characteristic relates to
 - 2 mis-pronunciations.

- 1 5. The method of claim 3, wherein the segment parameters relate to prosodic
 - 2 characteristics.

- 1 6. The method of claim 5, wherein the prosodic characteristics are selected from a
 - 2 group comprising pitch contour, spectral envelope, volume contour and phone
 - 3 durations.

1 7. The method of claim 6, wherein the prosodic characteristics are further selected
2 from a group comprising syllable accent, language accent, stress and emotion.

1 8. The method of claim 1, wherein blending the first TTS voice and the second TTS
2 voice further comprises extracting a prosodic characteristic from the LPC residual of the
3 first TTS voice and the LPC residual of the second TTS voice and interpolating between
4 the extracted prosodic characteristics.

1 9. The method of claim 8, wherein the prosodic characteristic is pitch, wherein the
2 interpolation of the extracted pitches from the first TTS voice and the second TTS voice
3 generates a new blended pitch.

1 10. A method of generating a synthetic voice, the method comprising:
2 receiving a user selection of a TTS voice and a voice characteristic; and
3 presenting the user with a new TTS voice comprising the selected TTS voice
4 blended with at least one other TTS voice to achieve the selected voice characteristics.

1 11. The method of claim 10, further comprising:
2 presenting the new TTS voice to the user for preview;
3 receiving user-selected adjustments; and
4 presenting a revised TTS voice to the user for preview according to the user-
5 selected adjustments.

1 12. The method of claim 10, wherein generating the new TTS voice further
2 comprises interpolating between corresponding segment parameters of the first TTS
3 voice and the at least one other TTS voice.

1 13. The method of claim 11, wherein the segment parameters relate to prosodic
2 characteristics.

1 14. The method of claim 13, wherein the prosodic characteristics are selected from a
2 group comprising pitch contour, spectral envelope, volume contour and phone
3 durations.

1 15. The method of claim 14, wherein the prosodic characteristics are further selected
2 from a group comprising: syllable accent, language accent, stress and emotion.

1 16. The method of claim 10, wherein the blended voice is generated by extracting a
2 prosodic characteristic from the LPC residual of the first TTS voice and the LPC residual
3 of the second TTS voice and interpolating between the extracted prosodic characteristics.

1 17. The method of claim 10, wherein the user-selected voice is blended with a
2 plurality of other TTS voices to generate the new TTS voice.

1 18. The method of claim 10, wherein the blended voice is generated by extracting a
2 prosodic characteristic from the LPC residual of the first TTS voice and the LPC residual
3 of the second TTS voice and interpolating between the extracted prosodic characteristics.

1 19. The method of claim 18, wherein the prosodic characteristic is pitch and wherein
2 the interpolation of the extracted pitches from the first TTS voice and the second TTS
3 voice generates a new blended pitch.

1 20. The method of claim 10, wherein the voice characteristic relates to mis-
2 pronunciations.

1 21. A system for generating a synthetic voice, the system comprising:
2 a module for presenting a user with a plurality of TTS voices to select at least one
3 voice characteristic;
4 a module for receiving a user-selected first TTS voice, a user-selected second
5 TTS voice, and at least one user-selected voice characteristic; and

6 a module for generating a new TTS voice by blending the first TTS voice and the
7 second TTS voice and according to the user-selected voice characteristic.

1 22. The system of claim 21, wherein the module that generates the new TTS voice
2 further interpolates between corresponding segment parameters of the first TTS voice
3 and the second TTS voice.

1 23. The system of claim 22, wherein the segment parameters relate to prosodic
2 characteristics.

1 24. The system of claim 23, wherein the prosodic characteristics are selected from a
2 group comprising pitch, contour, spectral envelope, volume contour and phone
3 durations.

1 25. The system of claim 24, wherein the prosodic characteristics are further selected
2 from a group comprising: syllable accent, language accent, stress and emotion.

1 26. The system of claim 21, wherein blending the first TTS voice and the second
2 TTS voice further comprises extracting a prosodic characteristic from the LPC residual
3 of the first TTS voice and the LPC residual of the second TTS voice and interpolating
4 between the extracted prosodic characteristics.

1 27. The system of claim 26, wherein the prosodic characteristic is pitch, wherein the
2 interpolation of the extracted pitches from the first TTS voice and the second TTS voice
3 generates a new blended pitch.

1 28. A method of generating a text-to-speech (TTS) voice generated by blending at
2 least two TTS voices, the method comprising:
3 establishing a voice profile for each of a plurality of TTS voices, each voice
4 profile having speaker-specific parameters;

5 receiving a request for a new TTS voice from a user; and
6 generating the new TTS voice by blending speaker-specific parameters obtained
7 from the voice profiles for at least two TTS voices.

1 29. The method of claim 28, wherein the speaker-specific parameters comprise at
2 least prosodic parameters associated with each TTS voice.

1 30. The method of claim 29, wherein the speaker-specific parameters further
2 comprise speaker-specific pronunciations.

1 31. The method of claim 28, wherein the speaker-specific parameters are related to at
2 least one of the group comprising: frame-based, phoneme-based, syllable-based and
3 general characteristics.

1 32. A text-to-speech (TTS) voice generated from a method of blending at least two
2 TTS voices, the method comprising:

3 establishing a voice profile for each of a plurality of TTS voices, each voice
4 profile having speaker-specific parameters;
5 receiving a request for a blended TTS voice from a user; and
6 generating the blended TTS voice by blending speaker-specific parameters
7 obtained from the voice profiles for at least two TTS voices.

1 33. The TTS voice of claim 32, wherein the speaker-specific parameters comprise at
2 least prosodic parameters associated with each TTS voice.

1 34. The TTS voice of claim 33, wherein the speaker-specific parameters further
2 comprise speaker-specific pronunciations.

- 1 35. The TTS voice of claim 34, wherein the speaker-specific parameters are related to
- 2 at least one of the group comprising: frame-based, phoneme-based, syllable-based and
- 3 general characteristics.